

# **Comparative and Functional Analysis of Cardiovascular-Related Genes**



**Len Pennacchio, Ph.D.  
Genome Sciences Department  
Lawrence Berkeley National Laboratory**

## **Sifting Comparative Sequence: Lipid Metabolism Genes and Regulation**

### **Outline**

#### **Computational Tools and Databases**

- VISTA
- Cardiovascular Gene Resource
  - Examples
- VistaBrowser
  - Human/Mouse Genome Comparison
  - GenomeVISTA

#### **Identification of a Novel Gene (ApoAV)**

- Functional Characterization

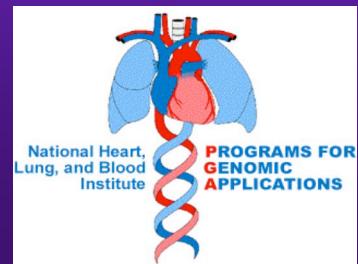
# PROGRAMS FOR GENOMIC APPLICATIONS

---

**MISSION:** To develop new resources, reagents, and education programs for investigators engaged in NHLBI-related research.

---

*National Heart, Lung, and Blood Institute (NHLBI)*  
National Institutes of Health



## NHLBI PGA Research Network

---



## The Human Genome: Feb. 2001



Today: 96% Finished    3% Draft

## Genomes with Significant Sequence Available



Mouse



Drosophila



Fugu



Rat



Ciona (Sea Squirt)



Zebrafish



Tetraodon

## Genomes in the Pipeline



Honeybee



Sea Urchin



Xenopus



Chimpanzee



Chicken



Dog

## Why compare the sequence of different organisms?



- Powerful in gene predictions
- Helps in understanding evolution
- Differences between evolutionary closely related species may explain biological differences
- Non-coding sequences conserved between species are reliable guides to regulatory elements

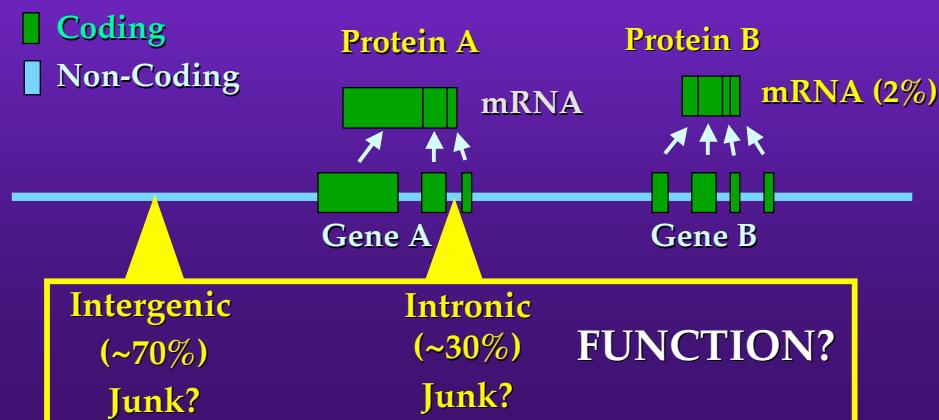
## Conserved Sequences are Functionally Important.



Find & Study these First

## Inventory of Mammalian DNA

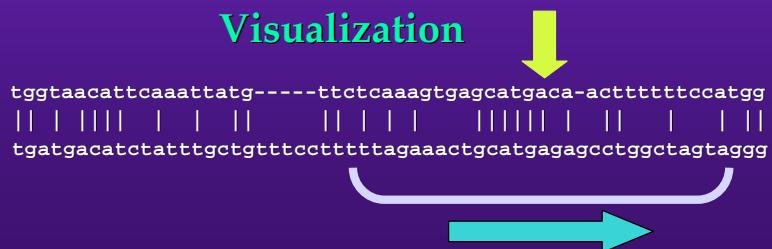
Coding ~2%  
Non-coding ~98%

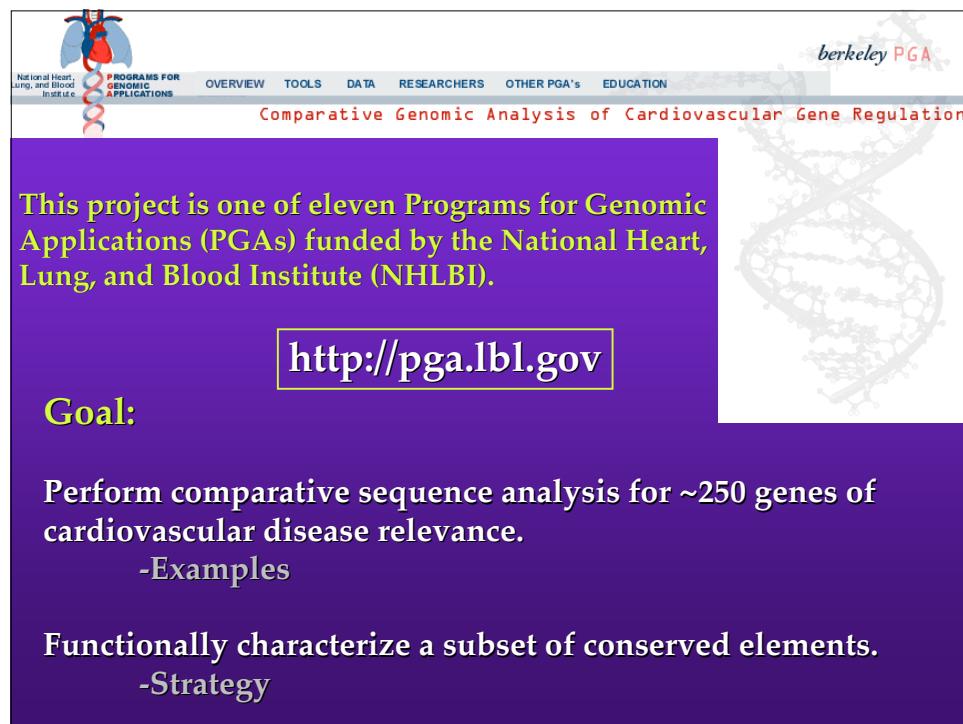
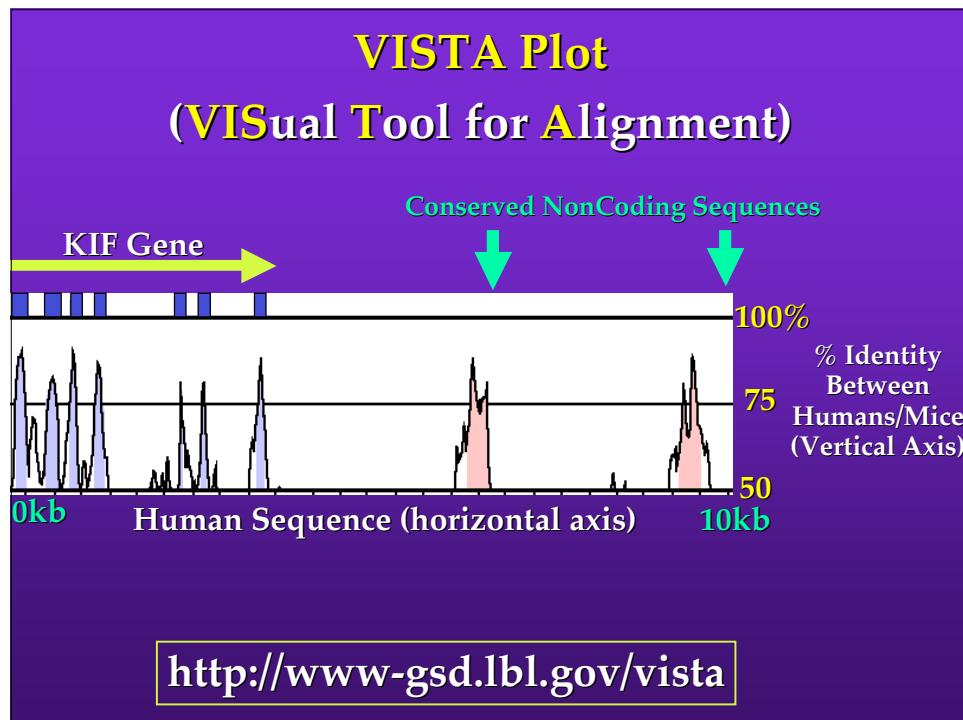




# AVID – The Alignment Engine Behind VISTA (Lior Pachter)

- **Very fast global alignment** of megabases of sequence.
  - **Provides details** about ordered and oriented contigs, and accurate placement in the finished sequence.
  - **Full integration** with repeat masking.





<http://pga.lbl.gov/>

*berkeley PGA*

**Cardiovascular Comparative Genomic Database (CVCGD)**



This database includes well-studied CV genes, for which an understanding of regulation should provide insights into CV relevant biological issues. While only a fraction of these genes will be characterized in the PGA biological projects over the 4-year time period of this program, the sequence of ~200 genomic intervals containing CV genes will be obtained and comparatively annotated and included in the CVCGD.

The database contains a variety of information for each gene relevant to this project:

- Gene name;
- Gene ID in the OMIM database (**OMIM**);
- Human map location (**HM**);
- GenBank accession number for human cDNA (**HC**);
- Mouse map location (**MM**);
- GenBank accession number for mouse cDNA (**MC**).

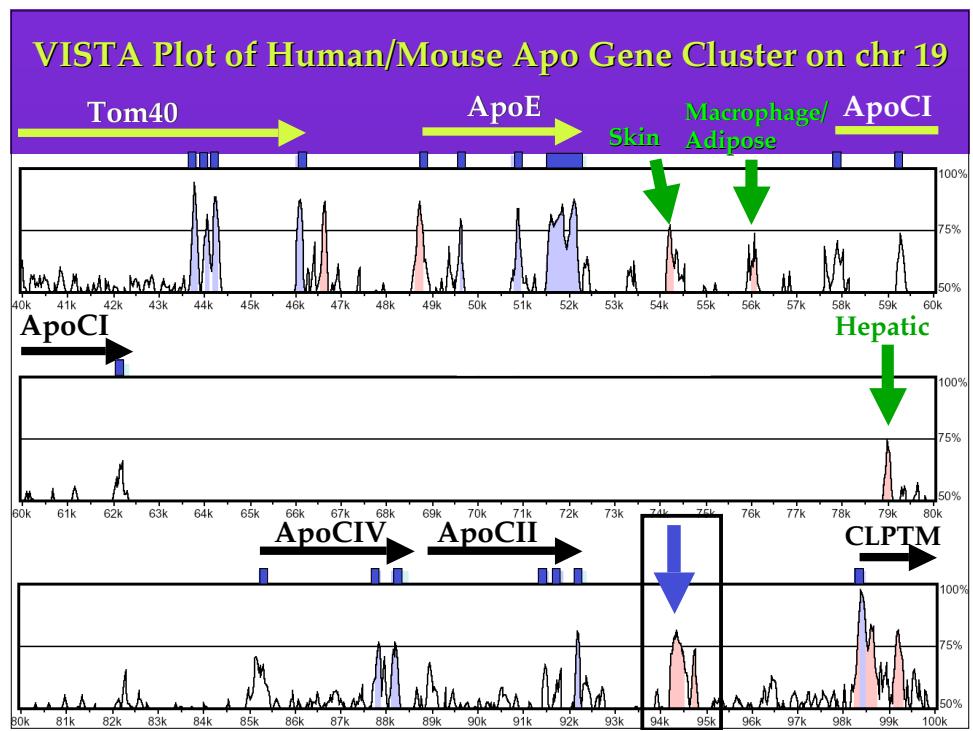
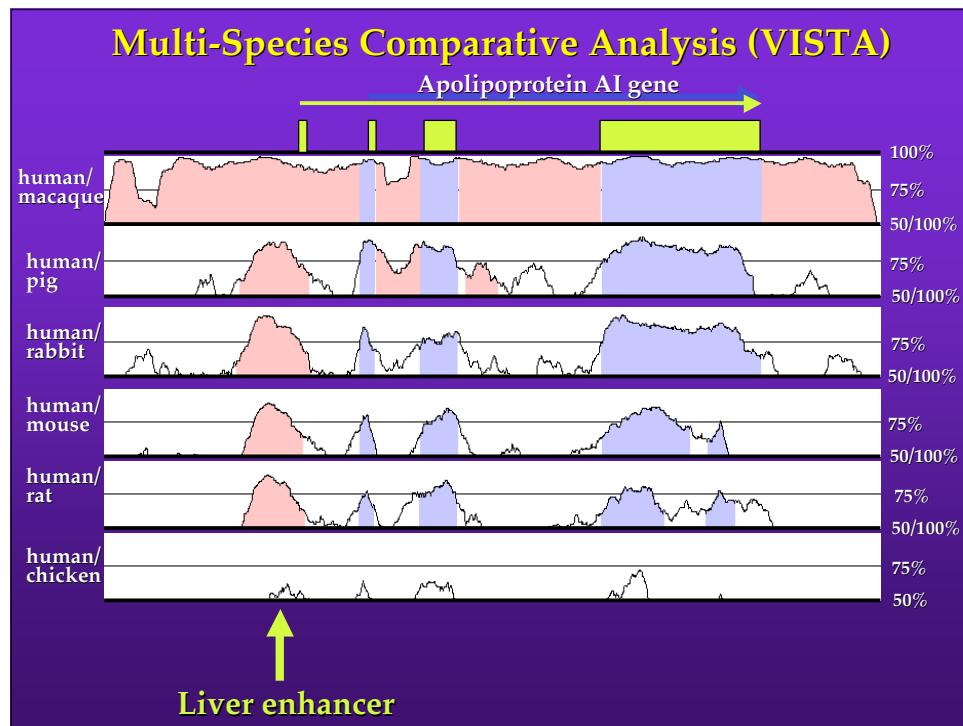
**SEARCH the CVCGD**

- [by gene name and abbreviation](#)
- [sorted alphabetically](#)
- [by categories](#) (groups of diseases).

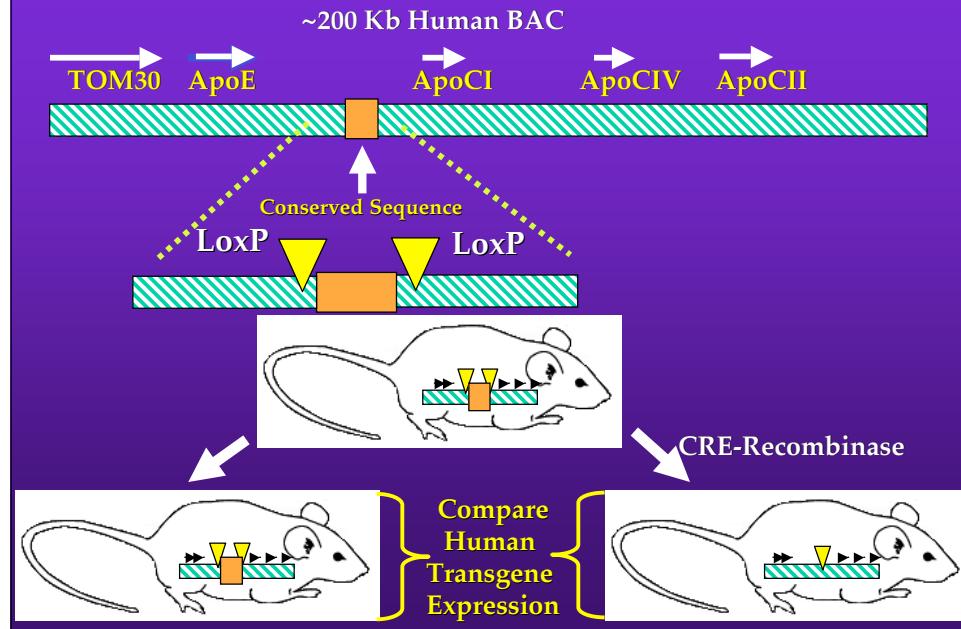
## Example

**Apolipoprotein A1 (APOA1)**

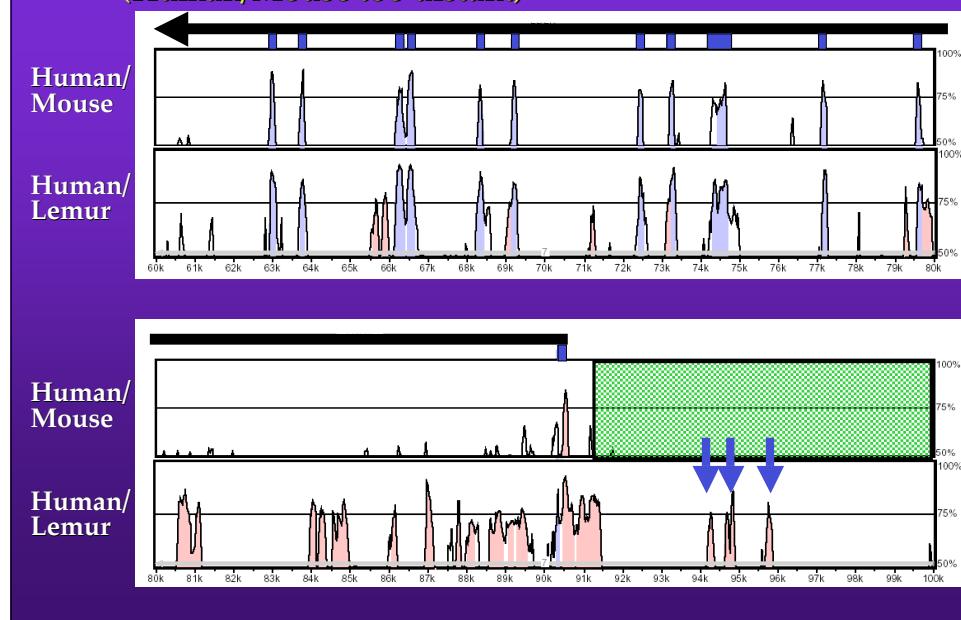
- Category: Atherosclerosis
- Gene ID in the OMIM database: [107680](#)
- Human map location: 11q23
- GenBank accession number for human cDNA: [NM\\_000039](#)
- Mouse map location: 9
- GenBank accession number for mouse cDNA: [NM\\_009692](#)
- [Annotation of the human sequence](#) (Explanation of annotation)
- Human-mouse alignment: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)
- [List of conserved regions](#) (human-mouse)
- Human-rabbit alignment: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)
- [List of conserved regions](#) (human-rabbit)
- Human-lemur alignment: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)
- [List of conserved regions](#) (human-lemur)
- Human-chimp alignment: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)
- [List of conserved regions](#) (human-chimp)
- Human-baboon alignment: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)
- [List of conserved regions](#) (human-baboon)
- Human-chicken alignment: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)
- [List of conserved regions](#) (human-chicken)
- All alignments displayed on the same scale: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)

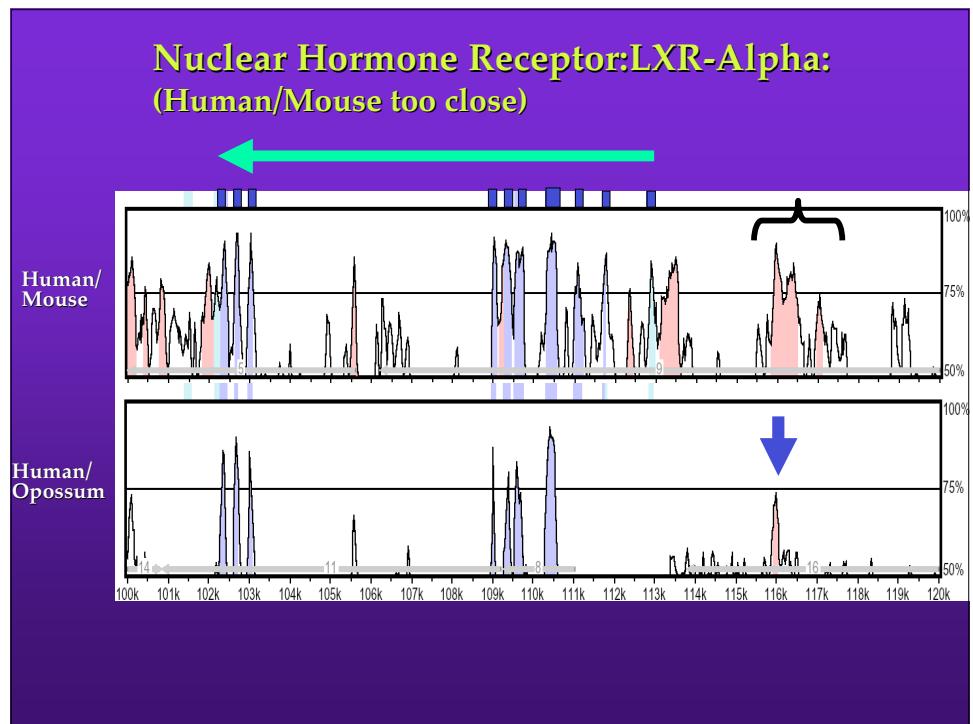


## Characterizing Conserved Sequence Function



## Low-Density Lipoprotein Receptor (LDLR): (Human/Mouse too distant)



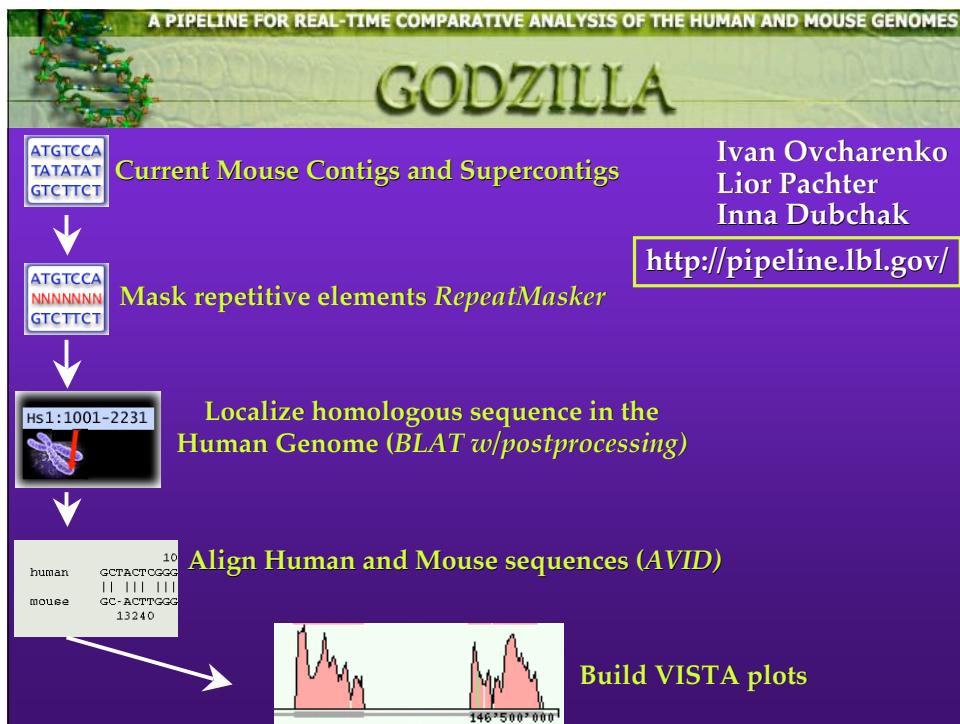
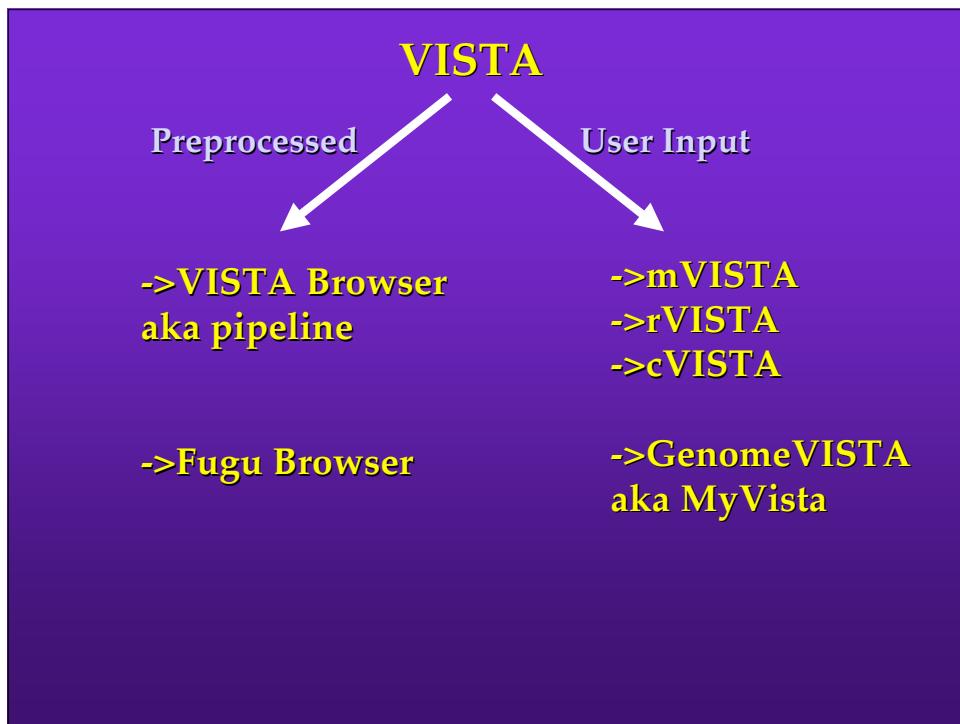


<http://www-gsd.lbl.gov/vista>

**mVISTA:** main VISTA  
-standard comparative sequence plots

**rVISTA:** regulatory VISTA  
-conserved transcription factor binding sites

**Availability:**  
Web-based  
Stand Alone Package



**VISTA Browser**

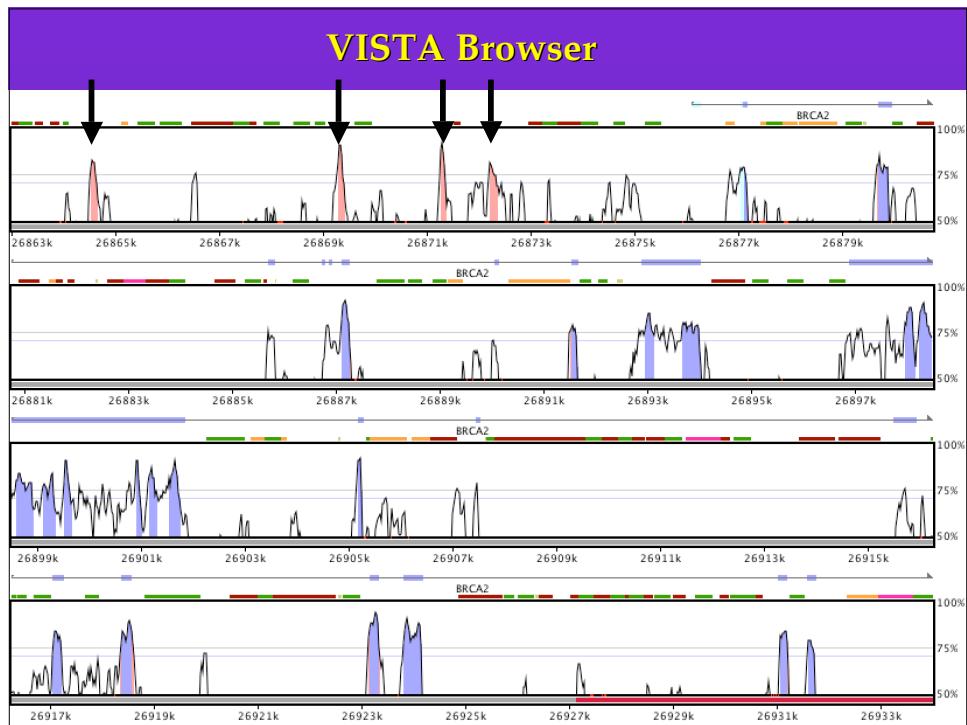
**Pre-processed Whole Human/Mouse Genome Comparison**

THE BERKELEY GENOME PIPELINE **CODZILLA**

Compare the Human and Mouse Genomes

Please enter a position in the Human Genome (June 2002 draft assembly), select the browser to display comparative analysis results, and press the submit button:

<http://pipeline.lbl.gov/>



**GenomeVISTA**

**Self-Input Sequence Comparison  
to Human, Mouse or Rat Reference Genomes**

**Submit a Request**

**Sequence**  
(choose one of the three options)

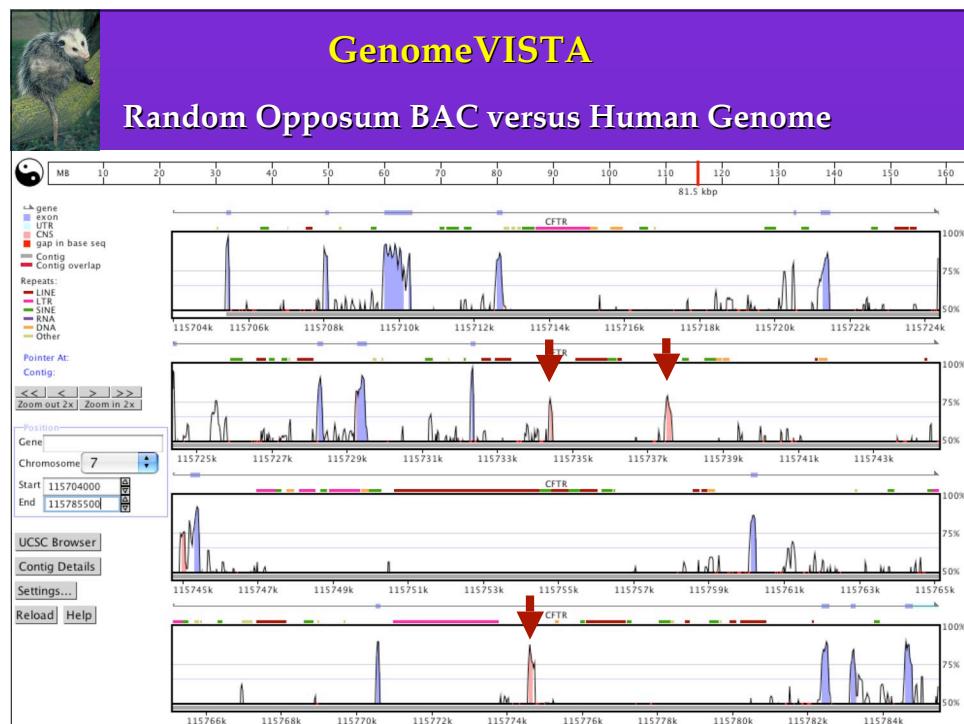
Paste a **Query Sequence** (FASTA format finished sequences only, 300K max)  
Draft sequences can be all entered at once, each config starting with > and the sequence name

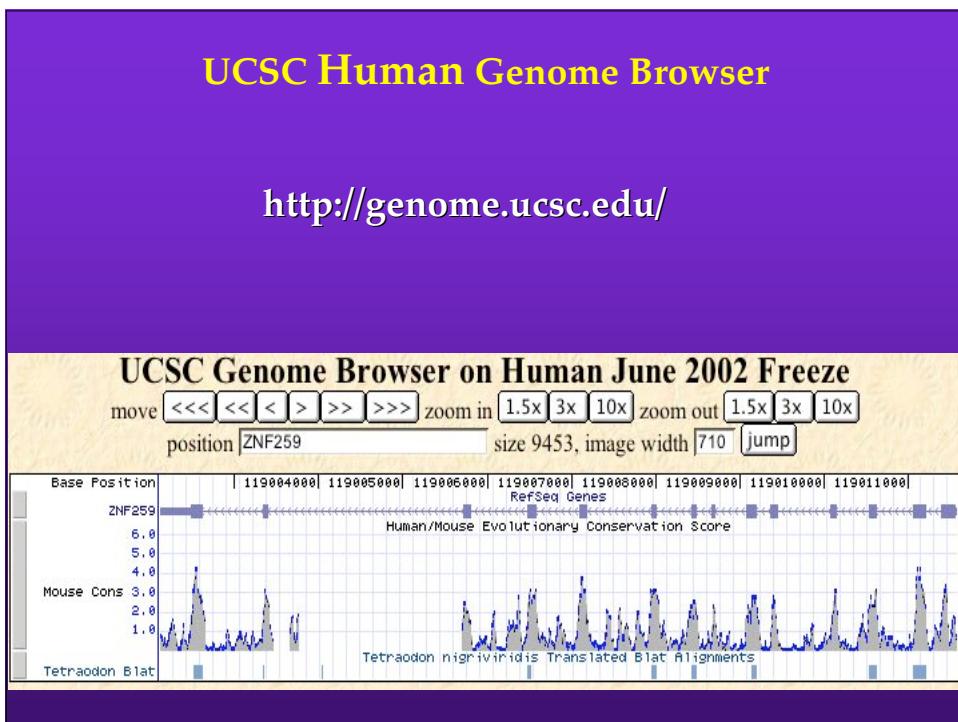
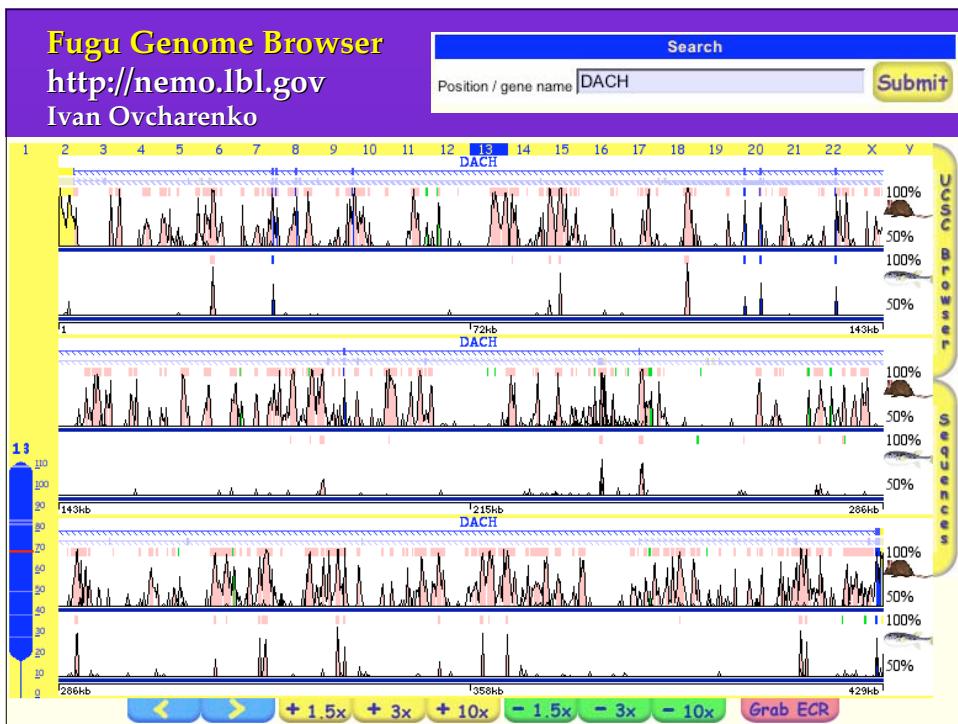
Alternatively, you can also select a file or enter a GenBank identification number:

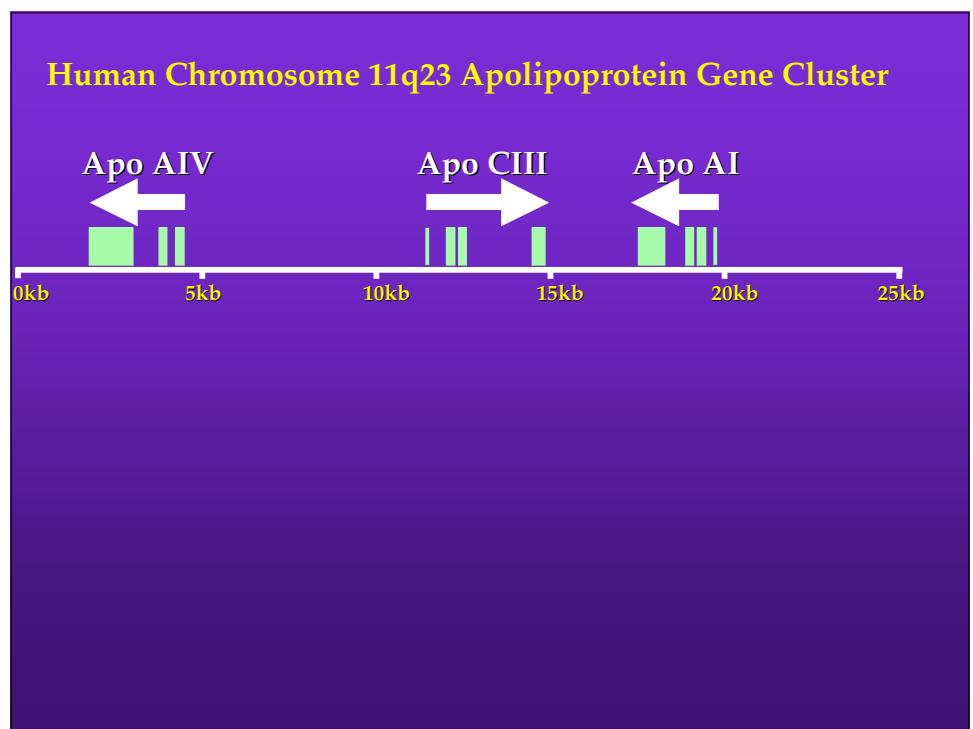
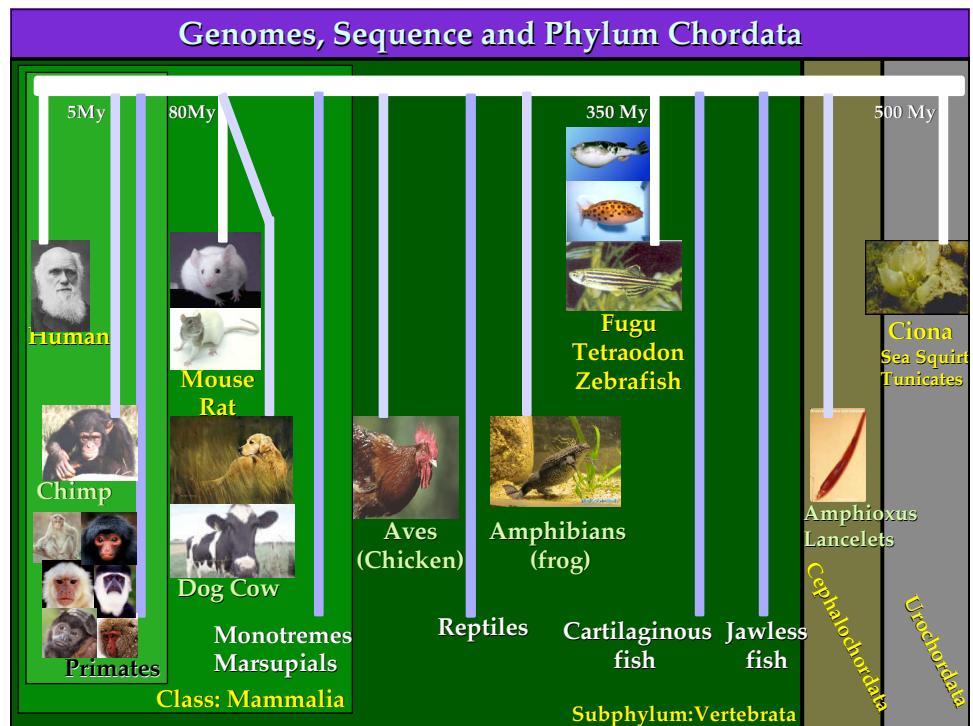
**FASTA**   Or **GenBank**   
Text files only. Word documents are not accepted. Sequences should be in FASTA format  
GenBank Locus:  
Accession or GI Number

 **opossum**

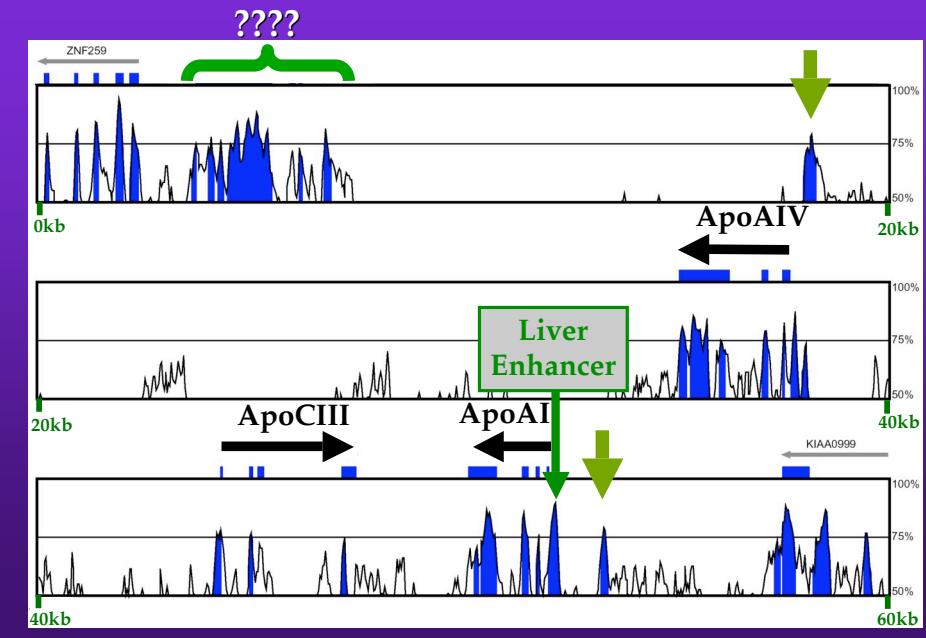
<http://pipeline.lbl.gov/>



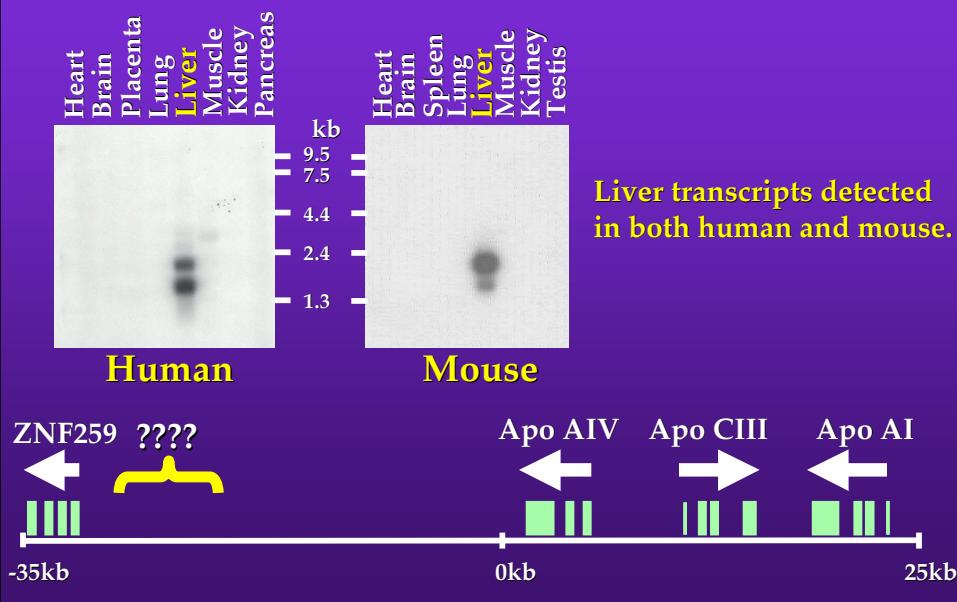




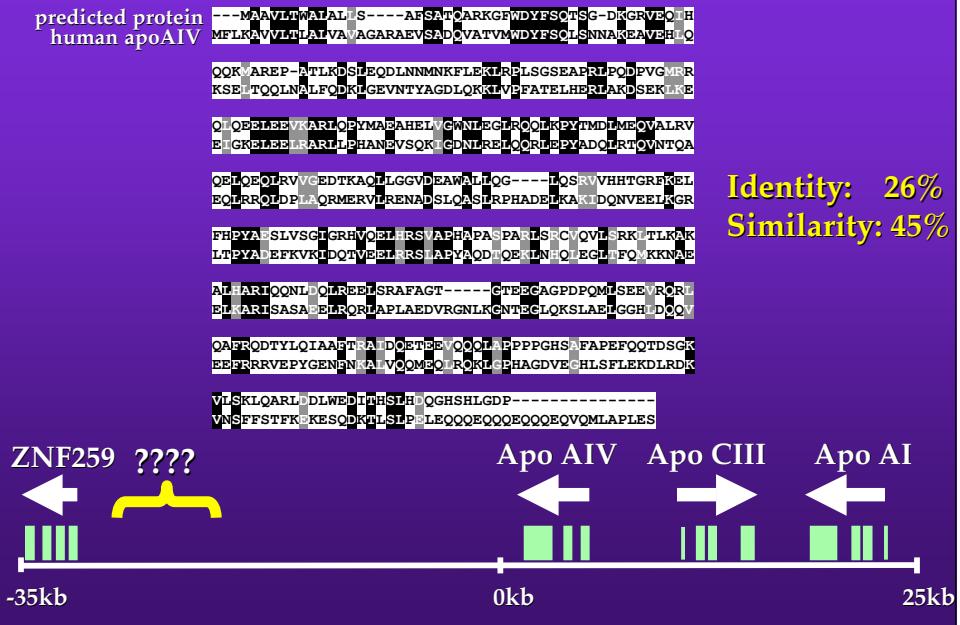
### Human/Mouse Apolipoprotein Gene Cluster Sequence Comparison



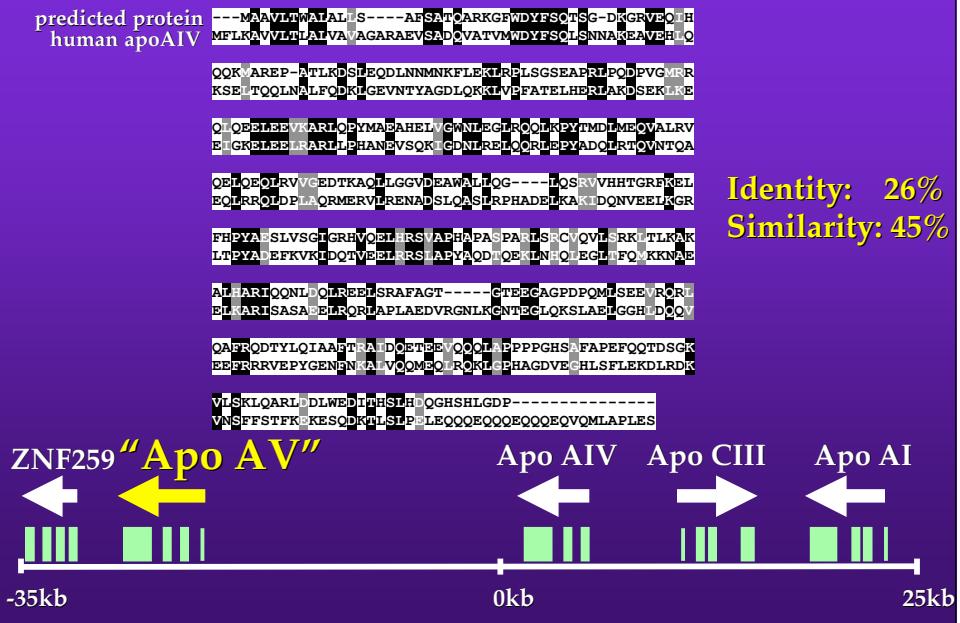
### Northern Blot Analysis of Conserved Sequence



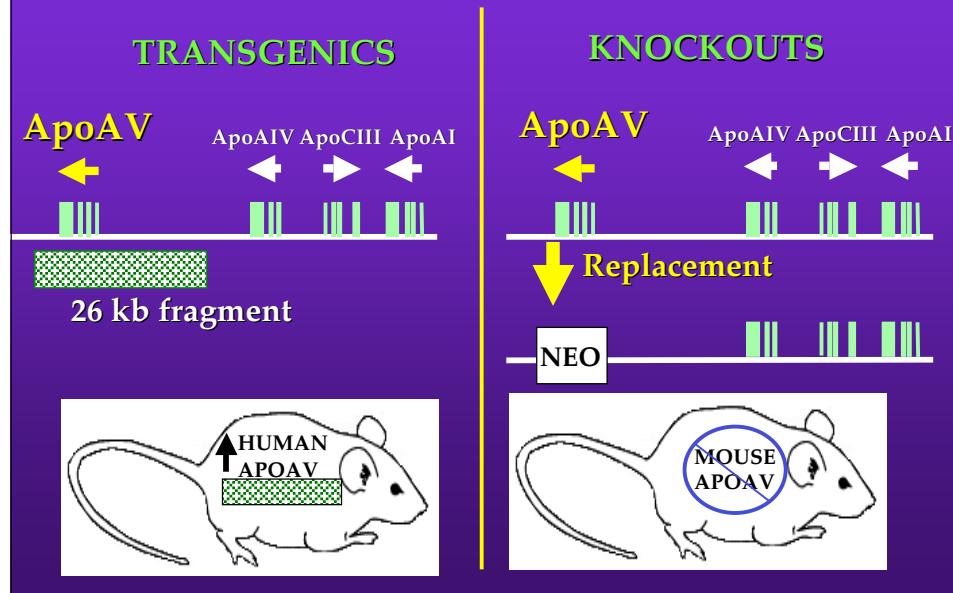
## Predicted protein has homology to ApoAIV



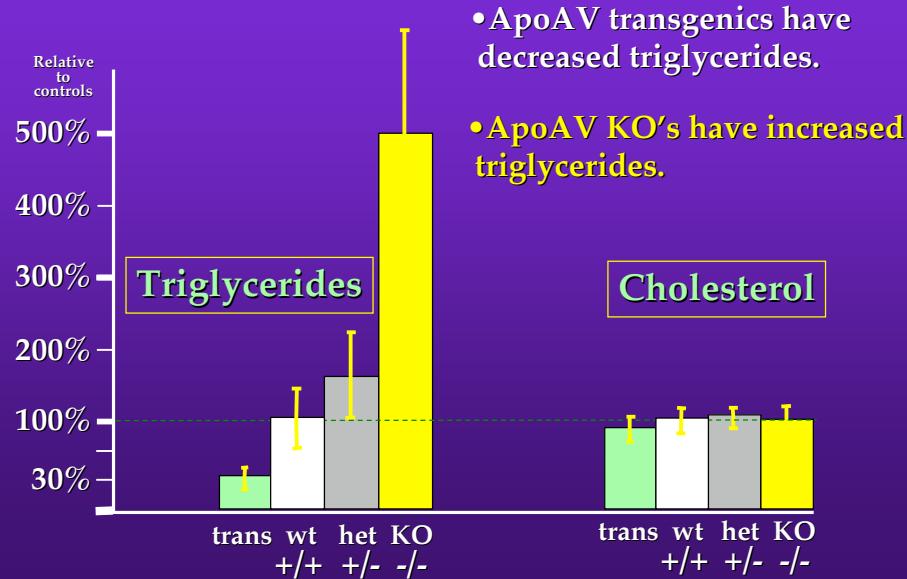
## Predicted protein has homology to ApoAIV



## ApoAV Mouse Studies



## ApoAV Transgenic and Knockout Plasma Levels



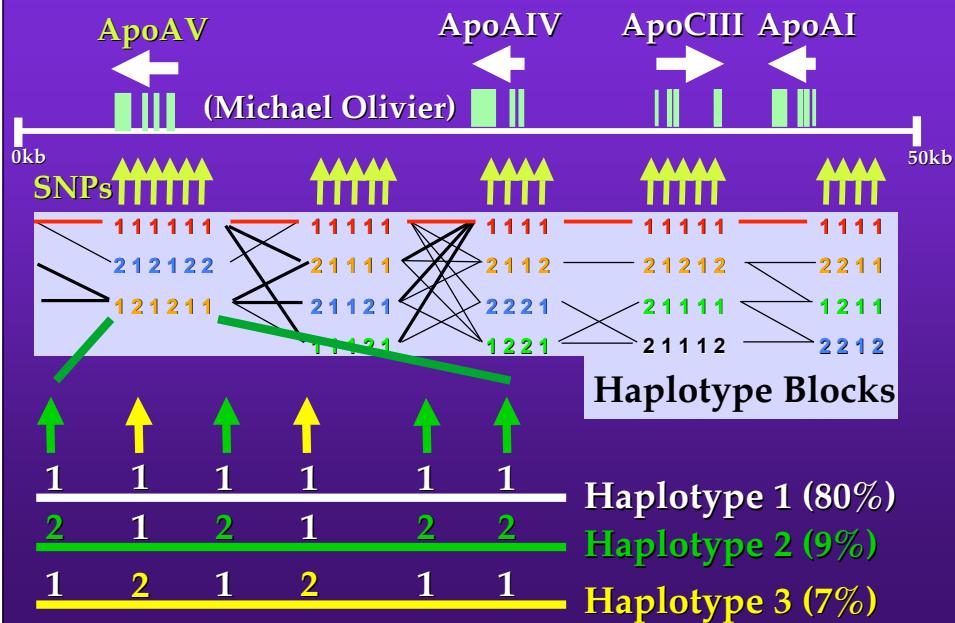
## Summary I: ApoAV

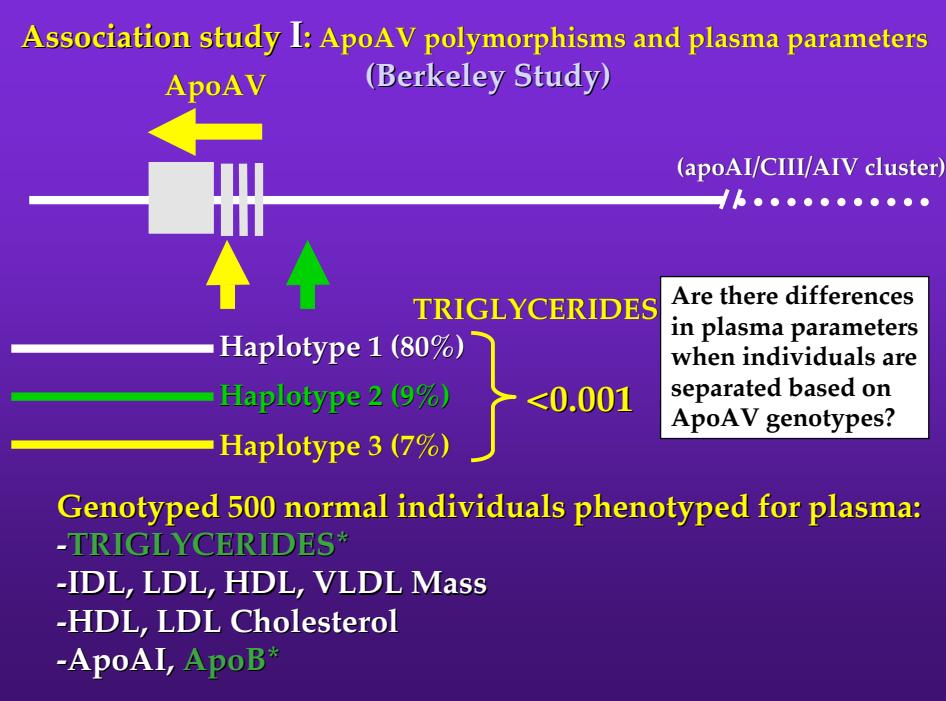
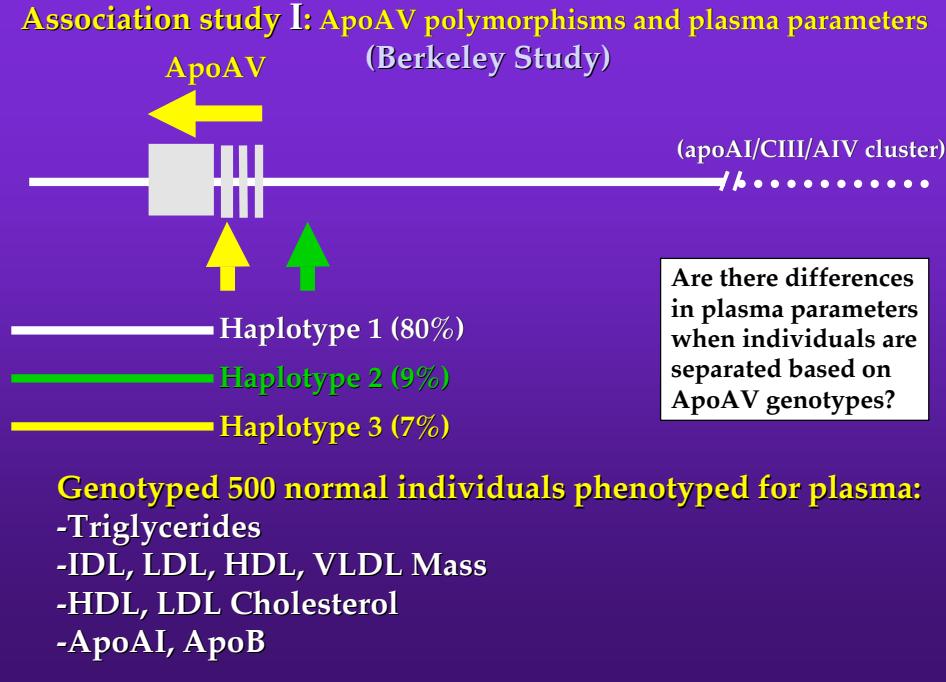
- A new apolipoprotein belonging to the ApoAI/CIII/AIV gene cluster.
- Expressed in the liver & associates with HDL/VLDL.
- An important modulator of triglycerides (TG) in mice.

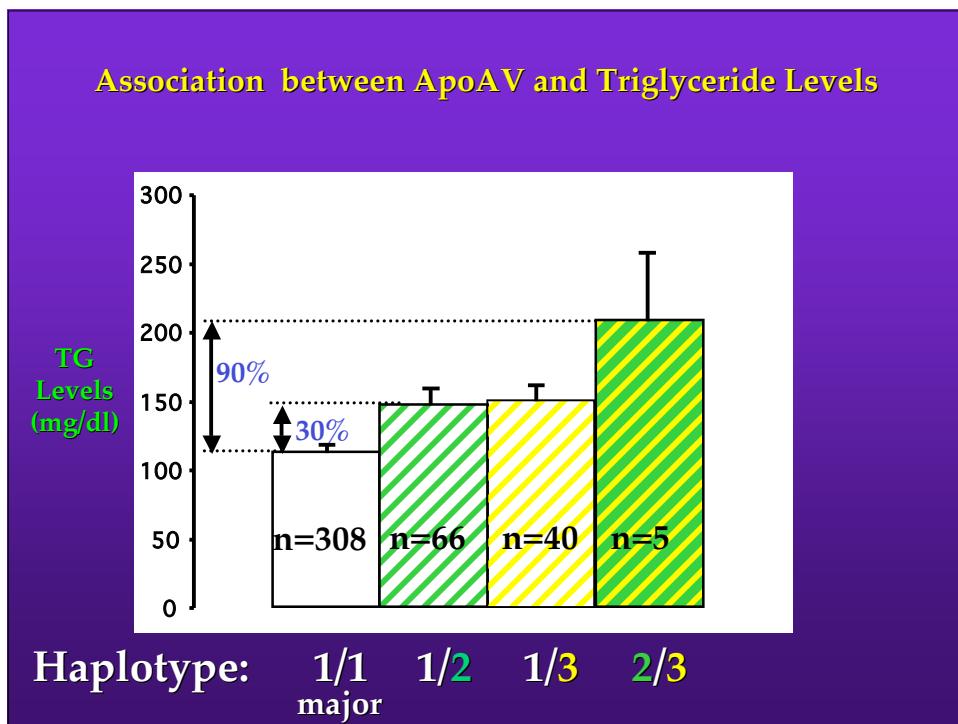
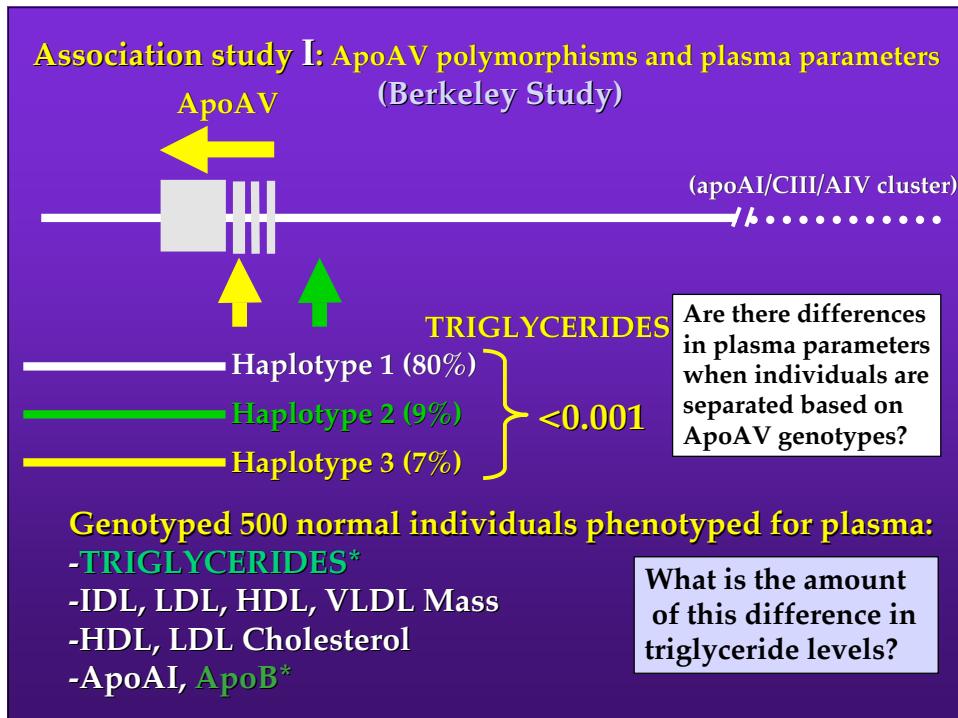


Is ApoAV involved in human biology/disease?

### SNP Identification/Haplotype Structure





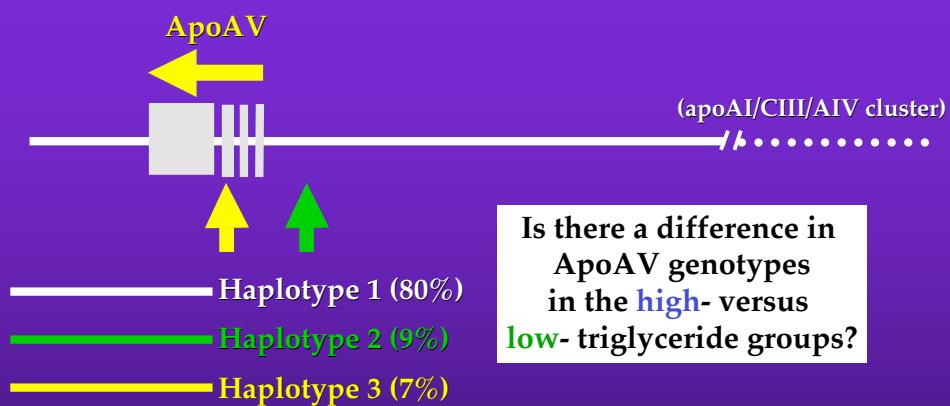


## Association Studies



Is this finding reproducible???

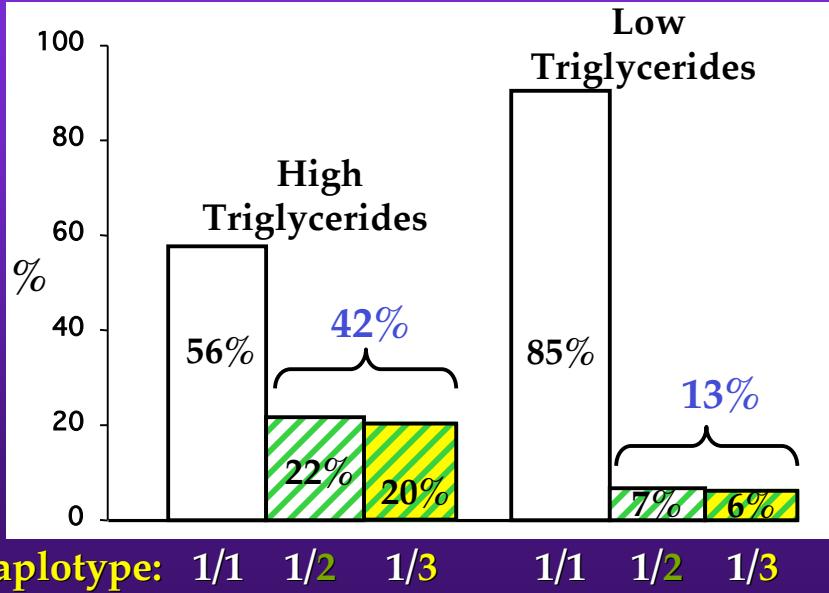
### Association study II: ApoAV polymorphisms and plasma parameters



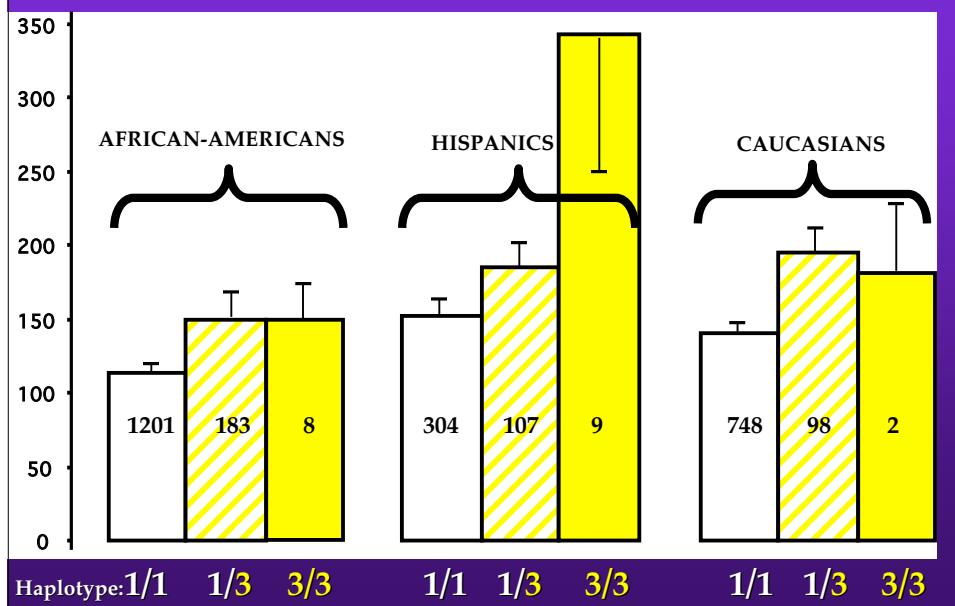
Genotyped 460 individuals stratified based on triglyceride levels.

High Triglycerides: Avg 340 mg/dl  
Low Triglycerides: Avg 50 mg/dl

**Association study II: ApoAV polymorphisms and plasma parameters  
(Dallas Study)**



**Association Study III: Dallas, TX**

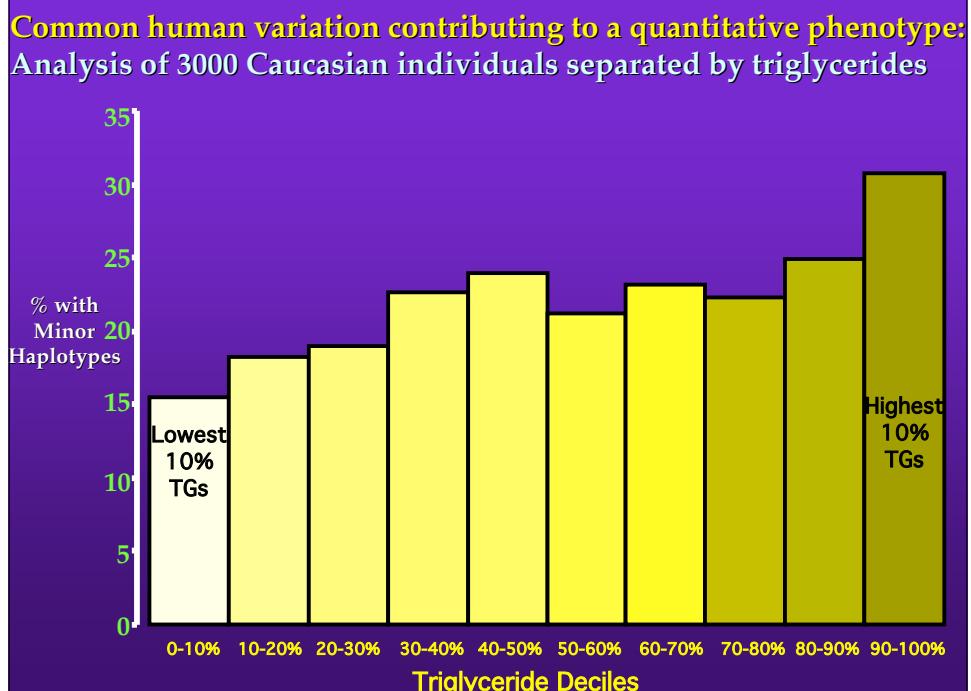


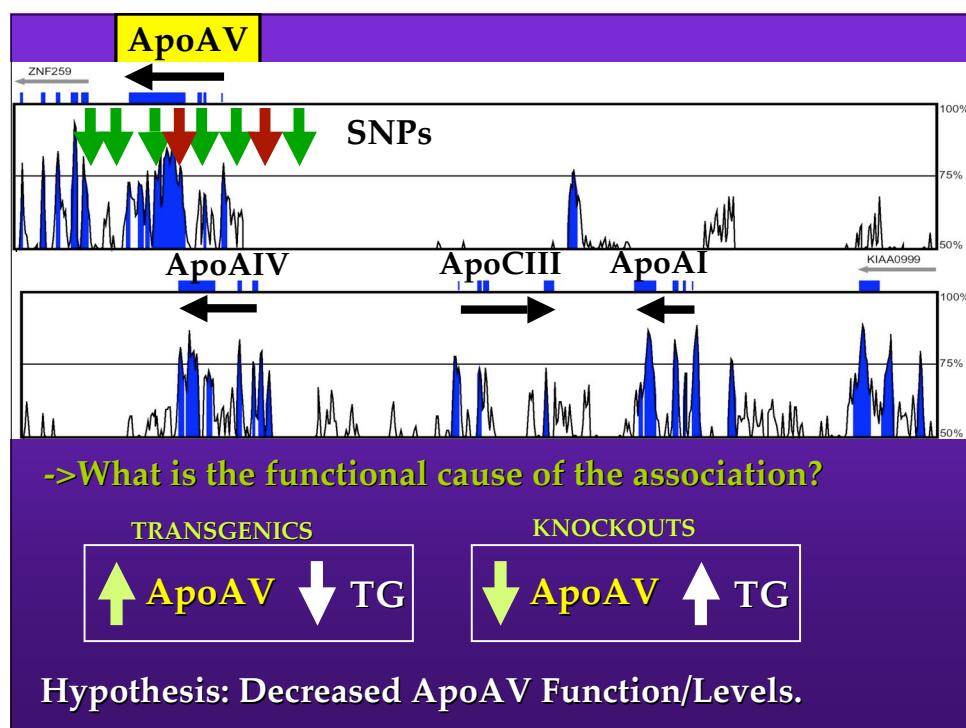
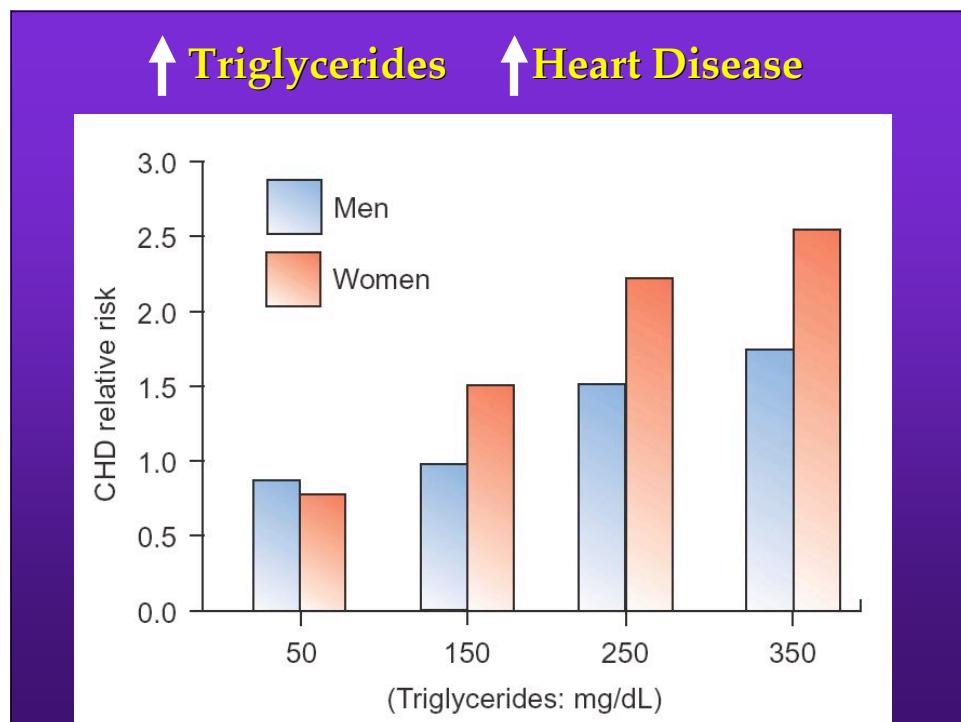
## ApoA5 and Triglyceride Levels

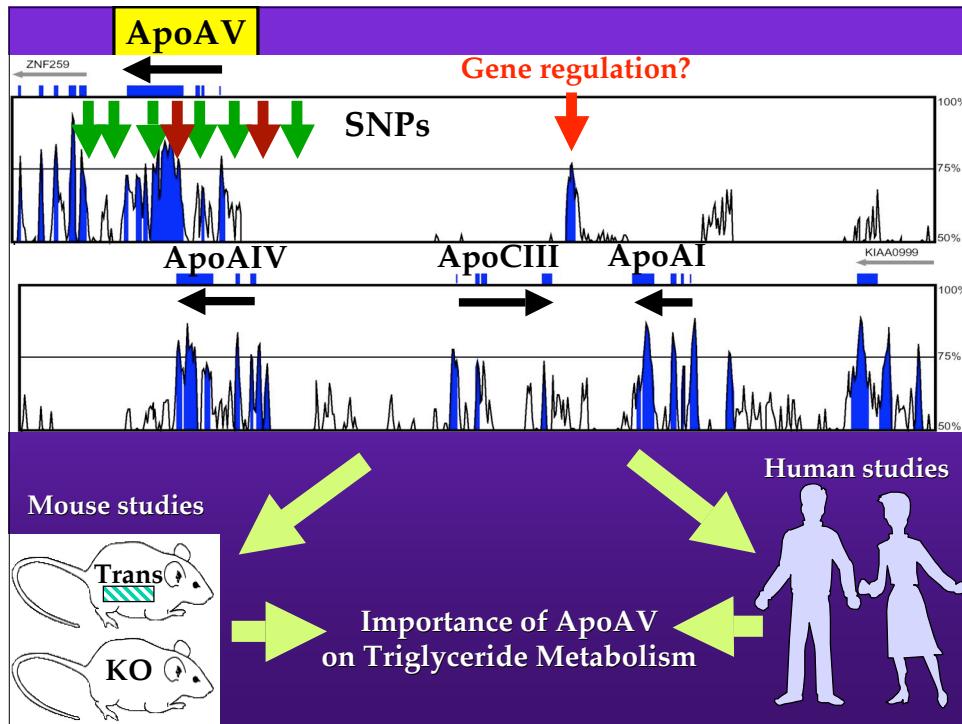
An example of common human variation contributing to a quantitative phenotype

Ethnicity:  
Caucasian  
African American  
Hispanic

Carriers of Minor Haplotype 2 and/or 3:  
24%  
36%  
51%







## Acknowledgements

### LBL

**Edward Rubin**  
Nadine Baroukh  
Elaine Gong  
Jennifer Akiyama  
Kathryn Houston  
Keith Lewis  
Willow Dean  
Jan-Fang Cheng  
Inna Dubchak  
Lior Pachter  
Ivan Ovcharenko  
Jody Schwartz  
Veena Afzal  
Xinli Yang  
**Ronald Krauss**  
Patricia Blanche  
Laura Holl  
Joseph Orr

### UT-SW

**Jonathan Cohen**  
**Helen Hobbs**  
Jaroslav Hubacek  
**Rayne Institute**  
Philippa Talmud  
Steve Humphries

### MCW

Michael Olivier

### NIH/NHLBI

<http://pga.lbl.gov>

### Pasteur Institute-Lille

Jamila Fruchart  
Jean-Charles Fruchart

### UCSF:

Brian Black

<http://www-gsd.lbl.gov/>